Pat Reynolds 4/9/98

The ICU rating system is based on processing events in sequence with ratings updates after each event: the order in which events are rated is significant. This is in contrast with the FIDE system where ratings are updated at 6-month intervals: here the sequence of events within the interval is irrelevant. One consequence of this is that while FIDE ratings 'turn over' in approx. 75 games for k=10 players, 50 games for k=15 & 30 games for k=30 - i.e. an established player with k=15 (below 2400) is effectively rated on his last 50 games - ratings in the ICU system can be slow to respond to recent form, certainly for established ('settled') players. Another consequence is that the ICU method is more difficult to maintain, makes more onerous demands on the rating officer & places an onus on tournament & league controllers to submit prompt & accurate results of games played to enable the system to function correctly. It is a moot point whether the ICU method can yield more accurate results than the FIDE method; in practical terms it is irrelevant because the ICU system has not been correctly maintained. It has been undermined by 3 factors-

1) Input
Garbage in, garbage out applies to any system. The rating officer does not receive prompt or accurate results of events played. He may not receive results at all. They may arrive months late. The format may be so sloppy that he cannot even establish who competed. He is unlikely to receive sufficient information to process new & provisionally-rated players correctly. Nor can he assume results submitted are correct and that byes/walkovers/defaults are so marked. Should he get through all of that & the only problem was a late report, the system has still been compromised - other later events have been processed out of sequence to meet demand for updates.

2) Software
The software used to maintain the system throughout the 90s is seriously defective. Among other things, it cannot calculate provisional ratings -a basic requirement- & has never done so.
Rather than debug this routine a patch was put in (early 90s) to treat new players as if they had played 2 I games & were now rated at 1000. In the mid-90s in response to UCU concerns about this the rating officer adopted a laborious method involving maintaining records outside the system for all provisionally-rated players, doing performance calcs separately, returning to the system to edit ratings for these players in order to process an event correctly. There are other issues. Errors once in the system are difficult & laborious to correct if it can be done at all – they have a ripple effect on subsequent events that cannot be fully corrected short of going all the way back & re-processing. See note 1.*
Tournament processing requires processing all rounds in an event for every player - e.g. if a player plays one game in a league division then necessary to enter 10 byes for the result to be accepted
Reporting is primitive and carries no date range options or useful filters- there is e.g. no simple method of producing a list of active players.
High-k players can earn bonuses but there are bugs in the bonus-calc routine. See Appendix 2.
Games processed under the 'Rate one Player' option can yield different results for high-k players than if entered as part of a tournament. Also this routine does not keep proper records & requires separate maintenance of external files to keep an audit trail.
The system is prone to crash.

3) Monitoring/Controls
It is unclear what controls if any have been applied to the system - at implementation or subsequently -to check that methods applied are consistent & statistically valid and to ensure the integrity of the rating list. For example- is the system intended to track / keep in synch with FIDE ratings? In the 80s this was the case & occasional adjustments were made to keep the two in line. There is evidence that ICU ratings are now quite deflated vs. FIDE - all recent lists show significant gaps between ICU & FIDE ratings for the same players. There are however significant FIDE interactions - results in representative events abroad are rated, also events where the player notifies the rating officer in advance; foreign players have FIDE ratings accepted here.

Comments

The theoretical basis for the current system is unclear - it appears to be something of an amalgam of the previous 2 systems & it is not evident that it would stand up to scrutiny even if the software worked.

The 'settled' criterion for a lower k is crude - based on a start date over 8 years ago but takes no account of inactivity over that time. The majority of players in the rating pool now come into this category -they have a lower k, cannot earn bonuses, and have no feedback mechanism applied to adjust for games against fast-improving provisionally-rated players & juniors. This is a likely cause of much of the deflation in the system.

The treatment of leagues is a major anomaly in a system that needs to rate events consecutively to function correctly. for rating purposes no event should go on longer than 3 months & the leagues should at a minimum be rated in 2 halves.

I have cited my own 97-8 results to June 1 as an example of other anomalies that can arise.
(Ref. ICO website, results, q1430.htm) - this shows 37 games played, score 50% against opposition averaging 1940, and a closing rating of 1873. 37 is a lot of games & one might expect a rating close to 1940 as a result. Also the last event was the Fide qualifier and a Fide performance of 1970 was less than break-even off 1877 ICU. One could argue from this alone that the ICU system is lacking both internal consistency & external validation.

The assimilation of foreign players also requires attention. Say a visiting 2500 player plays a few games at Kilkenny/Bunratty, & performs below par. The 2500 rating is taken as if established & the next list shows Bloggs IM 2470 near the top. What does this mean? The casual reader might take it that Bloggs had dominated ICU tournaments & performed well ahead of native players. Why not calculate his performance & show Bloggs 2300 based on 10 games?

Note 1*
There is a device re the current system to assist processing that has not been used to date. Rating officer tried correcting provisional ratings before processing an event but found it did not work & the original values were used. These are stored in temporary files & the Update Ratings routine does not refresh. The trick is to save tournament data to floppy prior to processing & backup the event from the floppy - the import will delete the files from floppy but refresh ratings. Can save a number of events in this way & later edit / make corrections as necessary & process the events in correct sequence. It is still quite onerous but just about possible to backtrack, correct errors in the system & reprocess subsequent events.

Recommendations

l) Input.
Get serious about this, regardless of any other decisions. Ban worst offenders from running ICU events, apply fines, whatever it takes. This crap has been going on for at least 15 years now & the rating task is too onerous to tolerate it.

2) Software.
Take current system, walk to nearest skip & dump it.

There are various options on how to proceed, including-
a) Take the USCF system out if storage & apply it on the basis that the software actually works, has a substantial track record & is statistically valid.
b) Buy/acquire rating software used in other countries
c) Decide on what is required & commission the software.
d) Adopt a simpler system - see Note 2* below

3) Monitoring/controls
Decide on the basis of the rating system & in particular if should track FIDE. If it should then implement whatever adjustments are necessary to bring them in line. As many Irish players now appear on both lists if may be sufficient to take a direct comparison as a basis for adjustment - check this & related queries with Fred Harte and/or Kevin O'Connell that the procedure is statistically valid.
If deflation in the ICU system is directly attributable to the current software then I imagine that adjustments should reflect level of activity over the past 7 years. Review periodically.

Appendix 3 contains rating information on active 1800+ players for reference.

Note 2*
This is an idea for a simpler system.
There is a database query tool called results.exe already in use on the ICO website which picks up individual results including reg. no., name & opponents rating, and in summary form calculates #won, #drawn, #lost & average strength of opposition. This is all the information required to calculate performance ratings. Use or adapt it to rate all active players on their last 50 games, 30 games for juniors. If this is a problem it would scarcely tax the IT skills of the executive to import the database files into a spreadsheet & do the same thing. See 4 for a hastily assembled example. Adjust the average opposition rating by applying the 350 limit on rating gaps. Impose a time limit of 5 years for inclusion of games - a rating list is not supposed to be a historical document. If player has less than 50 games on the system then take all their games. If less than 20 then provisional. Update the list just once a quarter - provisional ratings first - apply feedback mechanism as necessary (again consult FH/KOC for statistical validity) - then rate established players.

Systems

To 1989:
(Ref: Fiacla Fichille 1983-5 'Grading Matters' series)

K factors: k=32 <2000, k=16 >2000

Rating change calc:
+16 for win, -16 for loss; (8 for k=16)
discount rating gaps above 350;
add 4% of rating gap to lower-rated player, subtract 4% from higher-rated (2% for k=16)

Provisional ratings:
Calculate performance Rp = Opp_average + 400(#wins-#losses)/#games
Rating is weighted average of performances
Acceleration - if Rp>1200 then 1200 taken as new base rating
Rating established after 12 games

Bonuses/feedback:
Good performances earn bonuses, e.g. threshold 38 for 6 games, any gain above 38 doubled: +48 becomes +58 etc.
In processing an event provisional ratings done first, then a second pass using post-event ratings for these players; bonus earners ditto. Reasoning is that these players wrongly rated /underrated to begin with, feedback mechanism required to rate their opponents correctly.

Monitoring/controls:
Pool monitored to maintain integrity of system & control inflationary/deflationary effects (as in feedback mechanism above) so that e.g. a 1500 rating should indicate the same strength of player regardless of era. Drift control used for external validation - local pool aligned to Fide pool on grounds that Fide pool inherently more stable - interactions between strong established players - while local pool contains all sorts of players - young, old, casual, transient. Periodic checks that 2 systems in line by calculating performances in international events & adjustments made if necessary.

USCF: (2 yrs approx)
R = Ro + k (W-We) formula for rating change.
K factors: k=32 < 2100, k=24 for 2100-2399, k=16 > 2400.
Ratings provisional for first 20 games
Ratings <1000 treated as =1000 in rating opponents
Matches: maximum change ±50 per match, ±200 over 3 yrs from match play alone.

Current: (last 7 yrs approx):
The Timeout software product used is written in Clipper, a DBase compiler. Executables are chess.exe (3/4/92), cngrate.exe (20/7/90), chessind.exe (18/10/89) maintaining a number of database (.dbf) files.
Normal operation for established players as described by Stephen Morris - see Appendix 1.

R = Ro + k(W-We) used as in USCF system. We calculation can be reproduced accurately by applying the formula 1/(10^(r/400)+1) where r is the gap in ratings; 4 places of decimals are used.

Provisional ratings not calculated.
Different k factors used, notably a lower k for 'settled' players, higher k for juniors.
Bonuses applied for high-k players only - see Appendix 2.
No feedback mechanism.
No monitoring / controls.

Appendix 1

Rating Calculation by **Stephen Morris**

The Irish ratings were computerised in 1990 and the manner of calculating changes differs from the old system. This article describes the way in which ratings change and how to determine this manually without having to wait for the next rating list. This is of interest to the practical player.
Ratings are a system used to approximate playing strength and the combination of all the ratings in the system make up a pool. This pool is similar to the national money supply in that if ratings are too high then the system suffers from inflation (high prices) and if ratings are too low then deflation occurs (low prices).
The rating system strives to attain a happy medium where players have gradings which match their strength.
NOTE: The computer system uses four places of decimals but this article uses just two for simplicity. I would like to acknowledge John Crowley for kindly supplying me with information about the computerised rating system.

Method of Calculation:
Three items of information are needed for calculating rating changes:
1. K-factor
2. Expected score
3. Actual score
The K-factor is simply a number which reflects a player's age, rating and time at present rating. Young improving players have big K-factors (e.g. 40) while older, established players have smaller K-factors (e.g. 24).

Table 1 illustrates the different K-factors.

Player's Description	K-factor
Rating 2100 or higher	16
Player under 21 years of age	40
Player 21 years or older, rated <2100, and not settled	32 (It takes 8 years to settle)
All others	24

Table 1 K-factor Rules
The expected score is a number between 0 and 1 which reflects your chance of winning against a specific rating, e.g. I have an expected score of 0.25 against someone 200 points higher rated. My opponent has an expected score of 0.75
My expected score and my opponent's expected score both add up to 1.

The K-factor and the expected score are both obtained by consulting a lookup table. The actual score is the result of the game, 1, 0.5 or 0, i.e. a win, draw or a loss.
The rating change is calculated as follows:
Rating Change = K-factor * (Actual Score - Expected score).

This formula is the key to calculating ratings and the examples below illustrate its use in conjunction with Table 2.

Example 1
Player A is rated 1850 with a K-factor of 40 beats Player B rated 2250 with a K-factor of 16
The rating difference is 400 points. From Table 2 this rating difference gives player A an expected score of 0.10 and player B an expected score of 0.90
So, player A's rating increases by: K*(1.0 - 0.1) =40*(0.9) =36 points
Player B's rating decreases by: K*(0 - 0.9 = 16*(0.1) = 14 points

The Lookup Table
Table 2 below illustrates the expected scores for each player given a specific rating difference.

H: Higher rated player's expected score
L: Lower rated player's expected score

Rating Difference	H	L
0-6	0.5	0.5
7 -13	0.51	0.49
14-20	0.52	0.48

Etc.

Appendix 2

Bonus-calc routine
Formula: R = Ro +32(W-We)
if R>2099 (i.e. k=16) no bonus, exit
if k=24 no bonus, exit
increment counters- played, win, loss -with results
increment totopp - sum of opponents ratings
Average rating = totopp / #games
Rperf = avg rate+ 400(w-1) / #games ** Performance calc
n=#rounds-4
threshold=32+3n
if thresh >= gain no bonus, exit
if #rounds < 4 no bonus, exit ** (minimum 4 rds)

bonus=gain-threshold
R= R + bonus
if k=40 then R = Ro + (R-Ro)* 1.25 ** bug. Bonus value not updated

if R >Rperf then R = Rperf
if R >2099 then R=2099 ** bug. Too crude

Update, exit

Note that players on a 16 or 24 k cannot earn a bonus regardless of how well they perform in a tournament.

Routine contains 2 significant bugs -
a) bonus is calculated for k=32 and then adjusted for k=40 players, but the bonus value is not updated so that in the tournament reports the old rating + k(W-We) + bonus does not equal the new rating.

b) if the routine calculates a new rating >2099 it is set equal to 2099.

This is far too crude - e.g. a high-k player off 2095 could have an exceptional tournament result but pick up only 4 pts; with a moderately good result (i.e. no bonus) he could cross the 2100 threshold normally. Penalising a very good result can scarcely be the intent of the routine.